

WHAT IS CLAIMED IS:

1. A computer system for building large indexes,  
comprising:

5 an index engine operably configured for coupling  
with an indexer plug-in; and

an indexer plug-in having an index merger for  
concurrently merging sub-indexes created at a plurality of  
stages during indexing of content.

10

2. The system of claim 1 further comprising a gatherer  
engine operably coupled to the index engine for providing  
content to the index engine for indexing.

15 3. The system of claim 2 wherein the gatherer engine  
comprises a gatherer plug-in for selecting content for  
indexing.

4. The system of claim 2 wherein the gatherer engine  
20 comprises a content filter for extracting elements from  
content for indexing.

5. The system of claim 1 wherein the indexer comprises a content filter for extracting elements from content for indexing.

5 6. The system of claim 1 wherein the content comprises at least one member of the set comprising documents, images, audio streams and video streams.

7. The system of claim 1 further comprising a master  
10 index resulting from merging all of the sub-indexes created during indexing of content.

8. The system of claim 7 wherein the content index comprises a dictionary.

15

9. A computer readable medium having computer-executable components comprising the system of claim 1.

10. A method for building a large index in a computer  
20 system, comprising the steps of:

determining the number of stages for merging sub-indexes;  
determining the number of sub-indexes for each stage;  
building a sub-index in volatile memory;

storing the sub-index in persistent storage as belonging to one of the stages; and

merging sub-indexes at a stage before the number of sub-indexes at that stage exceeds the number of sub-indexes  
5 determined for that stage.

11. The method of claim 10 further comprising the step of determining to merge sub-indexes at a stage that has half of the number of sub-indexes determined for that stage.

10

12. The system of claim 10 further comprising the step of merging all sub-indexes to create a master index.

13. The method of claim 10 wherein the step of merging  
15 sub-indexes at a stage comprises storing the merged sub-index at the next stage.

14. The method of claim 10 wherein the step of  
determining the number of stages for merging sub-indexes  
20 comprises calculating the sum of half the number of sub-indexes at each stage and the product of the number of stages and the number of sub-indexes at each stage.

15. The method of claim 14 wherein the sum calculated is not greater than the number of persisted sub-indexes allowed for building the large index in the computer system.

5        16. The method of claim 10 wherein the step of determining the number of sub-indexes for each stage comprises calculating the sum of half the number of sub-indexes at each stage and the product of the number of stages and the number of sub-indexes at each stage.

10        17. The method of claim 16 wherein the sum calculated is not greater than the number of persisted sub-indexes allowed for building the large index in the computer system.

15        18. The method of claim 10 wherein the step of determining the number of stages for merging sub-indexes comprises calculating the product of the number of items for which index information may fit into an index in volatile memory and the quantity of half the number of sub-indexes at  
20 each stage raised to the power of one plus the number of stages for merging sub-indexes.

19. The method of claim 18 wherein the product calculated is not greater than the number of items to be indexed in the large index of the computer system.

5        20. The method of claim 10 wherein the step of determining the number of sub-indexes for each stage comprises calculating the product of the number of items for which index information may fit into an index in volatile memory and the quantity of half the number of sub-indexes at each stage  
10        raised to the power of one plus the number of stages for merging sub-indexes.

21. The method of claim 20 wherein the product calculated is not greater than the number of items to be  
15        indexed in the large index of the computer system.

22. The system of claim 10 wherein the step of merging sub-indexes at each stage comprises merging sub-indexes at each stage while continuing to index content.

20

23. The system of claim 10 wherein the step of merging sub-indexes at each stage comprises merging a copy of the sub-indexes for at least one stage.

24. The system of claim 10 wherein the step of storing the sub-index in persistent storage comprises storing the sub-index in persistent storage as belonging to a first stage.

5        25. The system of claim 10 further comprising the step of gathering content from the World Wide Web for indexing.

26. The system of claim 25 wherein the step of gathering content comprises gathering at least one member of the set  
10 comprising documents, images, audio streams and video streams.

27. The system of claim 10 wherein the step of merging sub-indexes at each stage comprises concurrently merging sub-indexes at different stages.

15

28. The system of claim 27 wherein the step of concurrently merging sub-indexes at different stages comprises using multiple processors.

20        29. The system of claim 10 wherein the step of building a sub-index comprises filtering information from an item being indexed.

30. The system of claim 29 wherein the step of filtering information comprises using a different filter for each different type of content.

5        31. A computer readable medium having computer-executable instructions for performing the method of claim 10.

32. A computer system for building a large index, comprising:

10        means for creating sub-indexes at different stages of a processing pipeline;

         means for concurrently merging sub-indexes at different stages of the processing pipeline; and

         means for continuously indexing content while merging the  
15 sub-indexes.

33. The system of claim 32 further comprising means for creating a master index after content has been indexed.

20        34. The system of claim 32 further comprising means for gathering content to index.

35. The system of claim 32 further comprising means for selecting content to index.

36. The system of claim 32 further comprising means for filtering information from content to build a sub-index.

5        37. The system of claim 32 wherein means for creating sub-indexes at different stages of a processing pipeline comprises means for determining the number of different stages of the processing pipeline.

10       38. The system of claim 32 wherein means for creating sub-indexes at different stages of a processing pipeline comprises means for determining the number of sub-indexes for each stage of the processing pipeline.

15       39. The system of claim 32 wherein means for concurrently merging sub-indexes at different stages of the processing pipeline comprises means for determining when to merge sub-indexes at different stages of the processing pipeline.

20

40. The system of claim 32 wherein means for continuously indexing content while merging the sub-indexes comprises means for adding new indexing information to sub-



indexes at different stages of the processing pipeline while sub-indexes are being merged.